

Introduction to PED Grade Score Comparison with CESE Canonical Scale Scores for 2012

"All calculations based on experience elsewhere fail in New Mexico" -- Lew Wallace, New Mexico Territorial Governor, 1878-81

The following contains a comparison of CESE scores with the PED's grade scores, among other comparison plots. The actual grades assigned are also shown, where appropriate. The same sets of plot types is shown for elementary, middle, and high schools. The CESE method uses only the results from the NM Standards Based Assessment (NMSBA) tests to see how the actual performance of a school compares to the predicted performance from the school's demographics. This allows one to subtract the predicted scores from actual scores to identify the significantly over-performing schools so that these can be studied for best practices to be used for other schools with similar demographics.

But one of the primary questions asked of CESE is how our scores compare with the PED's grade scores. To make this comparison, we must combine the math and reading scores that are output from the NMSBA test. We do this exactly the same way we do for the predicted scores in the CESE Method for Improving NM Schools. We use the well-verified mathematical canonical correlation technique. The state PED combines these scores by simple averaging. Ultimately, the results appear similar, but can vary significantly for any one given school.

CESE has decided to use only the canonically combined score. Initially, after the ABCDF Act was passed, we used all the variables called out in the statute: NMSBA grade translated to proficiency, growth, etc. The PED provided additional measures, such as dividing growth into two categories, including growth of lower performing student (lowest quartile) and the rest of the students. They also came up with other measures, some more subjective, such as "Opportunity to Learn (OTL)" which based its core on a survey of parents and attendance. So, there were a number of things used, some measured in percent proficiencies, some in scale scores (a common test measurement form from the NMSBA), and some in a random scale normalized to fit with the others. Additionally, the PED used "weightings" to combine these variables whose basis was apparently made on a "best guess" from the PED personnel and possibly others.

Interestingly enough, when CESE analyzed all these factors used by the PED individually, using actual data from the NMSBA, they generally were either correlated fairly well with the NMSBA scores, or they were sometimes negatively correlated. We also found that at least one of these major weighted factors (growth) was correlated so that it preferably gave some schools an advantage that scored on the disadvantaged end of the demographic spectrum compared to schools on the advantaged end of the demographic spectrum. We have nothing against helping the demographically disadvantaged, but think it inappropriate to do so artificially. That is unfair to the schools that receive an inflated score as well as to the schools that receive a deflated score.

The short version is that CESE decided to use the NMSBA Scale Scores only, and not try to account for all the other factors that were included. In fact, for the purposes of developing a method that shows schools **how** to grow academically, we could not justify the other factors. Additionally, the canonical correlation math weighted factors in accordance to their contributions. The other factors had very low impact and simply provided little information that was not

already contained in the NMSBA scores. Furthermore, we did not have access to student level data to work with and only used school level data. We do not believe this makes any significant difference but have no mathematical reasons to believe it would have based on the level of comparison we are performing.

Note that one of the questions frequently asked is whether or not the PED grades reflect school demographics. Thus far, the answer from the PED has been that the demographic effects have been removed or accounted for, quoting expert opinion, saying that using past data accounts for demographics (a bit of a simplification, but to the point). In reality, the PED grade scores are, indeed, correlated with NM school demographics, and plots have been included for each grade level (ES, MS, and HS) showing this. However, the PED's grading system has to follow federal guidelines in order to obtain the necessary waiver from restrictive NCLB requirements, resulting in information that does not show schools how to improve. These guidelines require that school performance as measured against the state standards be tested. To remove demographic effects is not allowable by the federal government because it would also remove the comparison to the standards. The CESE method actually shows both performance with respect to standards and the performance with the demographic effects removed. A state needs both to tell where a school stands, and how to improve. If the PED were to remove the effect of demographics among schools, the observation of the "Achievement Gap" and actual comparison to standards could not occur. And an exception would not be granted by the federal government to proceed with the ABCDF grading system. New Mexico would be back to looking at Annual Yearly Progress (AYP). Instead, CESE has a way of removing demographics that can be used as a basis for improving schools in NM, and need not replace the PED grade system. The PED grading system may not be perfect, but it is clearly correlated with schools' demographics, which is precisely what CESE uses to predict school performance. Thus, the two are generally well correlated with each other, which will be shown in the briefing that follows.

Finally, note that the school's performance as ranked by the PED is very close to the performance as ranked by CESE. It is not identical, but generally similar. Importantly, there are, however, some instances of significant differences of up to two grades using the PED grade scale), both for schools over-ranked and for schools under-ranked. Randomly looking at a few of these examples shows that this is almost certainly due to the arbitrary ranking factors added by the PED and possibly other influences. But no two cases we checked were identical. Generally, the differences appear to be caused by subjective items that may or may not be related to what is actually happening with a school's performance. We are not going to address this in any detail in the briefing that follows, but it is an important point – especially if your school happens to be one of the mis-ranked ones. This also begs the question as to why the correlations in the CESE and PED grades are fairly significant. The answer is actually pretty simple: most of the PED's scores, no matter what the weighting factors, trace back to the scale scores from the NMSBA. That is, they are already correlated. Once you have combined one or two, the rest are correlated such that little changes. And occasionally, you find where something unrelated (such as OTL) that simply combines with something else only remotely related that counters the actual performance outputs. When this happens, you get the grade swings that do not seem to make sense and show very low correlation with CESE combined scale scores.

So please review the plots in the briefing, and read the word slides, too. They have a tale to tell and should not be ignored.